Analysis of Fasciola cathepsin L5 by S2 subsite substitutions and determination of the P1–P4 specificity reveals an unusual preference

Luke J. Norbury a,1, Andrew Hung a, e, Simone Beckham b,2, Robert N. Pike b, Terry W. Spithill c, Charles S. Craik d, Youngchool Choe d, John V. Fecondo a, Peter M. Smooker a, e, *  

a School of Applied Sciences, RMIT University, Bundoora, Victoria 3083, Australia  
b Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia  
c Department of Agricultural Sciences and Centre for AgriBioscience, La Trobe University, Bundoora, Victoria 3083, Australia  
d School of Applied Sciences, RMIT University, Bundoora, Victoria 3083, Australia  
e Health Innovations Research Institute, RMIT University, P.O. Box 71, Bundoora, Victoria 3083, Australia

Abstract

Fasciola parasites (liver flukes) express numerous cathepsin L proteases that are believed to be involved in important functions related to host invasion and parasite survival. These proteases are evolutionarily divided into clades that are proposed to reflect their substrate specificity, most noticeably through the S2 subsite. Single amino acid substitutions to residues lining this site, including amino acid residue 69 (aa69; mature cathepsin L5 numbering) can have profound influences on subsite architecture and influence enzyme specificity. Variations at aa69 among known Fasciola cathepsin L proteases include leucine, tyrosine, tryptophan, phenylalanine and glycine. Other amino acids (cysteine, serine) might have been expected at this site due to codon usage as cathepsin L isoenzymes evolved, but C69 and S69 have not been observed. The introduction of L69C and L69S substitutions into FhCatL5 resulted in low overall activity indicating their expression provides no functional advantage, thus explaining the absence of such variants in Fasciola. An FhCatL5 L69F variant showed an increase in the ability to cleave substrates with P2 proline, indicating F69 variants expressed by the fluke would likely have this ability. An FhCatL2 Y69L variant showed a decreased acceptance of P2 proline, further highlighting the importance of Y69 for FhCatL2 P2 proline acceptance. Finally, the P1–P4 specificity of Fasciola cathepsin L5 was determined and, unexpectedly, aspartic acid was shown to be well accepted at P2, which is unique amongst Fasciola cathepsins examined to date.

© 2012 Elsevier Masson SAS. All rights reserved.
several clades that correspond to observed and predicted enzymatic activity [9,10].

The S$_2$ subsite exerts a dominant effect on the overall specificity of cysteine proteases [11]. Single amino acid substitutions lining this subsite are known to influence substrate specificity [12–14]. While FhCatL1 and FhCatL5 prefer aliphatic and Phe residues at the P$_2$ position and lack the ability to cleave substrates with a Pro residue at this important position [13–16], FhCatL2, FhCatL3 and FgCatL1G are able to cleave synthetic substrates with Pro residues at P$_2$ [12,13,15,17,18]. Studies indicate that this unusual ability can be partially attributed to the residue at position 69 of the mature enzyme, Y69 for FhCatL2, and W69 for FhCatL3 and FgCatL1G [12,13]. While a L69Y mutation was analysed in FhCatL5, the reciprocal mutation in FhCatL2 (i.e. Y69L) had not previously been examined.

By virtue of codon usage in the various isoenzymes, the presence of Fasciola cathepsin L proteases possessing other residues at aa69 would be expected to be present in Fasciola [9]. The generation of the Y69 codon requires 2 base substitutions from either the W69 or L69 codons; therefore other codons expected to be observed during the generation of these substitutions may include those encoding Phe, Cys and Ser. Since this prediction was made in 2003, cathepsin L proteases possessing F69 have been discovered in both Fasciola hepatica and Fasciola gigantica. Several F69 papain-like cysteine protease sequences have also been isolated from other trematodes, including Paragonimus westermannii and Schistosoma japonicum. A number of cysteine proteases have been discovered that possess a Ser residue at the position equivalent to position 69, but none with CatL-like homology have been isolated from trematodes, while only one C69 papain-like cysteine protease has been isolated, from Drosophila. This indicates that such CatL-like proteases may be structurally or functionally compromised.

Here, recombinant FhCatL5 and FhCatL2 with defined substitutions at aa69 were used to determine whether these substitutions result in active functional enzymes and their influence on substrate specificity of the active enzymes. These results have assisted the delineation of structural differences that alter substrate preferences among the broader cysteine protease family.

Furthermore, we analysed the complete P$_1$–P$_4$ specificity of the FhCatL5 enzyme using a positional scanning-synthetic combinatorial library, and compared the results to that previously obtained for other members of the family [14,19]. A striking difference in specificity at S$_2$ was observed. Homology modelling and computational docking calculations were employed to explore the molecular basis for the unusual capability of FhCatL5, amongst cathepsin L molecules, to accept Asp at the substrate P$_2$ position.

2. Materials and methods

2.1. Materials

The substrates Boc-Ala-Gly-Pro-Arg-AMC (AGPR), Boc-Asp-Pro-Arg-AMC (DPR), Z-Phe-Arg-AMC (FR), Bz-Phe-Val-Arg-AMC (FVR), Tosyl-Gly-Pro-Arg-AMC (GPR), Pro-Phe-Arg-AMC (PFR), Boc-Val-Leu-Lys-AMC (VLK), d-Val-Leu-Arg-AMC (VLR) and Boc-Val-Pro-Arg-AMC (VPR) were purchased from Sigma (Sydney, Australia). Oligonucleotides were obtained from GeneWorks (Adelaide, Australia).

2.2. Site-directed mutagenesis of FhCatL5 and FhCatL2

The expression plasmids encoding FhCatL5 (GenBank accession no. AF271385), FhCatL5 L69Y and FhCatL2 (U62289) have been previously described [13]. The aa69 substitutions in FhCatL5 and FhCatL2 were created using inverse PCR with Pfu polymerase using an annealing temperature of 52 °C, and an extension time of 17 min. Products containing substitutions to FhCatL5 were generated with primer combinations involving CL5reverse 5’ GCATACGAAATA TTGAAAACGATTTG and one of CL5C 5’ ATTTTCATACCCACCA CTAC; CL5$^{5'}$ ATTTTCATAGTCCACATAC or CL5F 5’ ATTTTC CATAAATCCACCATCC. These primers were designed to change the codon for Leu to that for Cys, Ser and Phe, respectively. The primers CL2reverse 5’ GCITATGAATITTTGAAACACAAAG and CL2L 5’ GTTT TCATCAATCTCCACCCG were used for the generation of FhCatL2 Y69L. The PCR products were treated with T4 polynucleotide kinase, then blunt end ligated. The sequence integrity of the constructs was confirmed by nucleotide sequence analysis.

2.3. Expression and purification of recombinant cathepsins in yeast

Transformations of plasmids into yeast and growth of the recombinant strains were as previously described [13]. The recombinant proteins were purified using nickel chelating and Mono S columns and processed at pH 4.5 in the presence of 10 mM cysteine as described previously [12]. Purified processed samples were stored in 100 mM sodium acetate, 1 mM EDTA, pH 5.5 (enzyme buffer) at −70 °C. The recombinant proteins were obtained at yields of approximately 3 mg/L. Some processing of the pro-protein occurred during preparation, as indicated by mature processed protein migrating below the unprocessed protein when analysed by SDS-PAGE (data not shown). Following the auto-activation step, only mature enzyme was visible when analysed by SDS-PAGE, and protease activity was confirmed for each recombinant protein by analysis with gelatin gel zymography undertaken essentially as described previously [20], and monitoring release of fluorogenic material against Z-FR-AMC (data not shown).

2.4. Enzyme assays of synthetic substrates

Purified mature enzyme was titrated against E–64 [21]. Protein concentration was determined using the modified Bradford assay [22]. Enzyme activity was assayed using the method described by [23] in enzyme buffer containing 10 mM cysteine at 37 °C. Kinetic constants were estimated by incubating the enzyme with varying concentrations of substrate, then plotting the rate of hydrolysis against substrate concentration. Curves were fitted using non-linear regression analysis with the program GraphPad Prism to yield estimates of $K_m$ and $V_{max}$.

Estimates of $k_{cat}/K_m$ were obtained by following substrate cleavage over 4 h at concentrations at least 10-fold lower than $K_m$. Assays were performed in enzyme buffer, pH 5.5, containing 10 mM cysteine in the presence of 4% (v/v) DMSO at 37 °C. Enzyme was incubated with 0.1–2 μM of substrate. Initial fluorescence ($I_0$), maximal fluorescence intensity ($I_{max}$) and the first order rate constant ($k$) where estimated by fitting a non-linear regression line to the fluorescence intensity (INT) against time ($t$) according to Eq. (1).

$$\text{INT} = I_0 + I_{max} \left(1 - e^{-kt}\right)$$  \hspace{1cm} (1)

When the initial substrate concentration is much less than the $K_m$, $k$ equals $k_{cat}/K_m$ [24]. Since $K_m$ was not determined for all substrates, the concentration of substrates used for the progress curves was determined by reducing the substrate concentration until the difference between $k_{cat}/K_m$ values was less than 10%.
2.5. P₁–P₄ specificity using a positional scanning-synthetic combinatorial library

The substrate specificities of FhCatL5 were determined using a complete diverse positional scanning-synthetic combinatorial library (PS-SCL) [19]. Screens were performed at 25 °C in 0.1 M sodium acetate, 0.1 M NaCl, 0.01 M DTT, 0.001 M EDTA, 0.01% (w/v) Brij-35, 1% (v/v) DMSO, pH 5.5. Aliquots of 25 nmol in 1 μl from each of 20 sub-libraries of the P₁, P₂, P₃, and P₄ libraries were added to the wells of a 96-well Microfluor-1 U-bottom plate (Dynex Technologies): the final concentration of each of the 8000 compounds per well was 31.25 nM. The assays were initiated by addition of pre-activated enzyme, and the reaction was monitored using a SpectraMax Gemini fluorescence spectrometer ( Molecular Devices) with excitation at 380 nm, emission at 460 nm and cutoff at 435 nm. Screens were performed in triplicate.

2.6. Protein/peptide modelling and automated docking

Homology modelling of FhCatL2 and FhCatL5 was undertaken using Modeller 9v2 [25] based on the crystal structure of F. hepatica procathepsin L1 (protein data bank (PDB) ID 2O6X [14]). Stereochemistry of the model was assessed using Procheck [26]. A ribbon representation of the protein is shown in Fig. 1A. To investigate the binding mechanisms of P₂ Asp tetrapeptides to the active site of cathepsin, docking studies (detailed below) were performed using the representative peptide Ac-QQDQ-ACC (corresponding to P₄–P₃–P₂–P₁, with ACC (7-amino-4-carbamoylmethylcoumarin) being the fluorophore), built using Deepview [27], as shown in Fig. 1B. ACC corresponds to position P₁. This sequence was chosen based on the P₁–P₄ specificity profiles of FhCatL5 (see below) and FhCatL2 [14], which indicate that Gln is well accepted (>50% of maximum rate) at the P₁, P₃ and P₄ positions.

Automated dockings of the tetrapeptide to FhCatL2 and FhCatL5 were performed using Autodock 4.2 [28]. Probe atom interaction energies were computed using 100 × 100 × 100 Å³ side grids with a spacing of 0.375 Å, with the grid box centred at C25 for both proteins. The ligand was subsequently docked to the proteins using the Lamarckian genetic algorithm according to the protocol and parameters detailed in Morris et al. [28], with the following exceptions: population sizes of 300, 25,000,000 evaluations and 270,000 generations were employed for all dockings. All rotatable torsion angles for the ligand, except peptide bond ω angles, were allowed to rotate freely, resulting in 22 rotatable torsions. The protein atoms were kept rigid. A total of 3000 docking runs were performed for each protein.

3. Results

3.1. Substrate cleavage assays

Fasciola cathepsin L proteases vary at several residues lining the S₂ subsite (Table 1). Most variation is observed at residues 69 and 209, both sites known to influence substrate specificity. Hydrolytic activity of FhCatL2 and FhCatL5 and their variants were examined against various substrates to determine the influence of the substitutions at aa69. Kinetic parameters against the substrates Z-FR-AMC and Tosyl-GPR-AMC were initially determined. Results coincided with those of previous studies [13–15]. FhCatL5 showed a clear preference for Z-FR-AMC over Tosyl-GPR-AMC, and although FhCatL2 was less efficient at cleaving Z-FR-AMC, it also readily cleaved Tosyl-GPR-AMC. In this study, FhCatL2 was almost equally efficient against Z-FR-AMC and Tosyl-GPR-AMC (Table 2, represented graphically in Supplementary Fig. S1).

The introduction of the L69S and L69C substitutions into FhCatL5 resulted in higher affinity for Z-FR-AMC, but the overall efficiency (k₉cat/Kₘ) for cleavage of Z-FR-AMC was reduced about 5-fold: the affinity for and efficiency for cleavage of Tosyl-GPR-AMC was also reduced. In contrast, the L69F and L69Y substitutions lead to a 2–3-fold increased efficiency towards Tosyl-GPR-AMC, although levels were still well below those observed for FhCatL2.

The introduction of the Y69L substitution in FhCatL2 resulted in a reduced affinity for both substrates, with the overall efficiency reduced somewhat more against Tosyl-GPR-AMC.

Analysis of progress curves against a wider range of substrates corroborated indications that FhCatL5 has a clear preference for Leu at P₂ and also confirmed the preference of the enzyme for a P₂ Phe over Pro (Fig. 2). The L69Y and L69W substitutions altered the S₂ subsite profile, resulting in an increased preference for Pro at P₂, although k₉cat/Kₘ levels were still several-fold lower than for Phe, while the efficiency against the aliphatic P₂ possessing substrates o-VLR-AMC, Boc-VLK-AMC and Bz-FVR-AMC all decreased. L69S and L69C substitutions resulted in generally decreased efficiency against most substrates tested (some were slightly higher, for example PFR-AMC), and efficiency against Boc-VLK-AMC remained comparable to FhCatL5.

The introduction of Y69L into FhCatL2 resulted in an increased efficiency against o-VLR-AMC, Boc-VLK-AMC and Bz-FVR-AMC, while the substrates containing Pro and Phe at the P₂ position were less efficiently catalysed.

3.2. P1–P4 positional scanning library for FhCatL5

The specificity of FhCatL5 was determined using a complete diverse PS-SCL (Fig. 3). The P1 preference of FhCatL5 was similar to that reported for FhCatL1 and FhCatL2 [14], as well as other parasite papain-like cysteine proteases and human cathepsins [19]. While this subsite of FhCatL5 readily accepted Gin, Glu and Arg residues, unlike FhCatL1 and FhCatL2, it had a clear preference for Lys residues. It also possessed lower relative specificity against Nle and Met residues compared to FhCatL1 and FhCatL2 [14].

The P2 specificity was also similar to that reported for FhCatL1 and FhCatL2 (excluding that seen for Pro residues with the latter enzyme), with a preference for aliphatic residues, the most favoured of which were Leu residues. Also similar to FhCatL1 and FhCatL2, despite showing efficient cleavage of substrates possessing P2 Phe, this amino acid was not identified as being particularly optimal at this site, while the low acceptance of Phe was expected. A significant difference at this site compared to other Fasciola cathepsins and papain-like cysteine proteases in general is the preference for Asp residues. This distinguishes it from the specificity of the other Fasciola cathepsins analysed to date. Modest acceptance of Asp residues at P2 has been shown previously for human cathepsin F (HuCatF) [19], although the preference was not as strong as shown for FhCatL5, while a Trypanosoma brucei cathepsin B has also been shown to accept P2 Asp residues [30]. Computational docking was employed to explore the molecular basis for the unusual preference of FhCatL5 for P2 Asp, as elaborated below.

The specificity for P3 and P4 residues displayed by FhCatL5 was fairly broad, although there were several subtle differences from FhCatL1 and FhCatL2, with FhCatL5 having a narrower specificity. The specificity at P3 for FhCatL5 was similar to that at P1, with pronounced specificity towards polar residues, with a noticeable preference for basic amino acids (Lys, Arg), while also accepting Ala and Pro residues. Distinct from FhCatL1 and, to a lesser degree, FhCatL2, aromatic and aliphatic residues were not acceptable at this site and neither were Gly residues. FhCatL5 shows the broadest specificity for P4 residues of any of the sites analysed, with a profile very similar to HuCatL, HuCatV and cruzain [19]. This site has a preference for polar residues, while aromatic and aliphatic residues were less readily accepted.

3.3. Modelling of substrate binding

Predicted docking poses of Ac-QQDQ-ACC to both cathepsin L subtypes using Autodock [31] were identified which most closely resemble the experimentally-acquired co-crystal structure of human cathepsin L and the peptide QLA (PDB ID 3K24 [32]), shown in Fig. 4. We specify two criteria which must simultaneously be satisfied in order for a docking pose to be deemed qualitatively similar to the co-crystal structure: 1) The peptide P2 sidechain lies within 6.5 Å of aa163; and 2) The peptide P4 backbone carbonyl lies within 6.5 Å of aa25. Based on the assumption that the H3 peptide binding configuration in this structure is representative of the proteolytically active conformation, this enabled determination of the extent to which FhCatL5 and FhCatL2 are capable of binding the tetrapeptide in configurations which are amenable to proteolytic cleavage. Additionally, it allows insights into the peptide–cathepsin L interactions which (dis)favour P2 Asp acceptance.

Amongst the 3000 predicted docking poses, five ligand-binding configurations that resemble that of the crystal structure exist for FhCatL5, while six were identified for FhCatL2. These predicted binding poses are hereafter referred to as “potentially proteolytic”

Table 1

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Genbank Acc. no.</th>
<th>S2 subsite preferences</th>
<th>S2 subsite amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>FhCatL1</td>
<td>AAB41670</td>
<td>Aliphatic &gt; Aromatic*</td>
<td>69* 70* 135* 160* 161 163* 209*</td>
</tr>
<tr>
<td>FhCatL2</td>
<td>AAF76330</td>
<td>Aliphatic &gt; Aromatic, Asp</td>
<td>Leu Met Ala Val Asn Ala Leu</td>
</tr>
<tr>
<td>FhCatL5</td>
<td>AAC47721</td>
<td>Aliphatic &gt; Aromatic, Pro*</td>
<td>Leu Met Ala Leu Asn Gly Leu</td>
</tr>
<tr>
<td>FgCatL1H</td>
<td>AAR08900</td>
<td>Unknown</td>
<td>Tyr Met Ala Leu Thr Ala Leu</td>
</tr>
<tr>
<td>FhCatL6</td>
<td>ARZ04042</td>
<td>Unknown</td>
<td>Gly Met Ala Met Asn Ala Ala</td>
</tr>
<tr>
<td>FhCatL3/FgCatL1G</td>
<td>ARZ0398/AAL23917</td>
<td>Gly, Pro, Aliphatic &gt; Aromaticbc,d</td>
<td>Trp Met Ala Val Thr Ala Val</td>
</tr>
</tbody>
</table>

Where * denotes residues whose sidechain extends into the subsite.

a Results obtained from Ref. [14].

b Results obtained from Ref. [11].

c Results obtained from Ref. [17].

d Results obtained from Ref. [29].

Table 2

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>$k_m$ (µM)</th>
<th>$k_m$</th>
<th>$k_{cat}$ (s⁻¹)</th>
<th>$k_{cat}$/$k_m$ (M⁻¹ s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FhCatL5</td>
<td>Z-FR-AMC</td>
<td>24.81</td>
<td>3.73</td>
<td>17.89</td>
<td>1.19 7.21 × 10⁴</td>
</tr>
<tr>
<td>FhCatL5 L69Y</td>
<td>tosyl-GPR-AMC</td>
<td>53.92</td>
<td>5.73</td>
<td>0.84</td>
<td>0.06 1.55 × 10⁶</td>
</tr>
<tr>
<td>FhCatL5 L69C</td>
<td>Z-FR-AMC</td>
<td>32.87</td>
<td>6.81</td>
<td>21.52</td>
<td>2.11 6.55 × 10⁷</td>
</tr>
<tr>
<td>FhCatL5 L69S</td>
<td>tosyl-GPR-AMC</td>
<td>63.46</td>
<td>5.49</td>
<td>2.54</td>
<td>0.09 4.00 × 10⁶</td>
</tr>
<tr>
<td>FhCatL5 L69F</td>
<td>Z-FR-AMC</td>
<td>13.72</td>
<td>1.52</td>
<td>1.92</td>
<td>0.08 1.40 × 10⁷</td>
</tr>
<tr>
<td>FhCatL5 L69S</td>
<td>tosyl-GPR-AMC</td>
<td>76.74</td>
<td>11.50</td>
<td>0.19</td>
<td>0.01 2.50 × 10⁸</td>
</tr>
<tr>
<td>FhCatL5 L69S</td>
<td>Z-FR-AMC</td>
<td>12.72</td>
<td>2.03</td>
<td>1.73</td>
<td>0.10 1.36 × 10⁸</td>
</tr>
<tr>
<td>FhCatL5 L69S</td>
<td>tosyl-GPR-AMC</td>
<td>126.90</td>
<td>16.06</td>
<td>2.07</td>
<td>0.27 2.10 × 10⁹</td>
</tr>
<tr>
<td>FhCatL5 L69S</td>
<td>Z-FR-AMC</td>
<td>19.11</td>
<td>2.51</td>
<td>17.93</td>
<td>0.87 9.38 × 10⁷</td>
</tr>
<tr>
<td>FhCatL5 L69S</td>
<td>tosyl-GPR-AMC</td>
<td>51.93</td>
<td>5.47</td>
<td>1.78</td>
<td>0.07 3.44 × 10⁵</td>
</tr>
<tr>
<td>FhCatL2</td>
<td>Z-FR-AMC</td>
<td>22.57</td>
<td>4.51</td>
<td>5.17</td>
<td>0.50 2.29 × 10⁵</td>
</tr>
<tr>
<td>FhCatL2</td>
<td>tosyl-GPR-AMC</td>
<td>36.15</td>
<td>5.65</td>
<td>7.08</td>
<td>0.57 1.96 × 10⁵</td>
</tr>
<tr>
<td>FhCatL2 V69L</td>
<td>Z-FR-AMC</td>
<td>50.08</td>
<td>5.65</td>
<td>5.22</td>
<td>0.36 1.04 × 10⁵</td>
</tr>
<tr>
<td>FhCatL2 V69L</td>
<td>tosyl-GPR-AMC</td>
<td>54.71</td>
<td>7.57</td>
<td>2.19</td>
<td>0.13 4.01 × 10⁵</td>
</tr>
</tbody>
</table>

Individual $k_m$, $k_{cat}$ and $k_{cat}/k_m$ values predicted from the non-linear regression fitting of Michaelis–Menten curves for cleavage of peptide substrates by cathepsins at pH 5.5.
poses. While all six of the identified proteolytic dockings for FhCatL2 share similar characteristics (discussed below), the dockings for FhCatL5 exhibit more structural diversity. However, one FhCatL2 structure (shown in Fig. 5A and B) was identified which corresponds more closely to the experimentally acquired structure than all others (including those of FhCatL2). Hereafter, we compare features of the “best” predicted dockings for FhCatL5 and FhCatL2 (Fig. 5) with those of PDB 3K24.

For both FhCatL5 and FhCatL2, the side chains of P1 and P3 (P1 (A) and P3 (Q) for HuCatL; P1 and P3 (Q) for FhCatL5 and FhCatL2) orient away from the active site cleft, and make no specific interactions, in agreement with 3K24. However, an important difference between the two subtypes is manifested in the orientation of the P1–P1’ carbonyl C–NH bond axis. For FhCatL5, this bond is oriented such that the P1 carbonyl carbon points towards H162, while the P1’ (ACC) peptide backbone N atom points away from H162. This is shown in Fig. 5A (circled) and 5B. This orientation is qualitatively similar to that of 3K24 (Fig. 4A and B, note that the C-terminal NH2 is not visible in the crystal structure; we have built it in for the purpose of illustrating the position of the P1’ peptide backbone N). We note that the carbonyl O of P1 makes direct hydrogen bonding contacts with H162 in 3K24. For FhCatL5, our best predicted docking indicates that P1 is within 6 Å of H162; nevertheless, the correct orientation of the P1 carbonyl O suggests that this ligand conformation is close to the proteolytically active state. In contrast, for FhCatL2, all “experimentally similar” predicted dockings indicate that the P1 carbonyl C–N bond axis is oriented incorrectly, such that the C points away from H162, while the N atom is more proximal to H162 (Fig. 5C and D (circled)). The P1–P1’ backbone orientation is stabilised by interactions with residues close to the catalytic dyad (C25–H162). For FhCatL5, correct orientation of the P1–P1’ backbone is stabilised by ring-stacking interaction between the ACC ring and W184 (Fig. 5A). For FhCatL2, the P1’ NH group hydrogen bonds with the backbone of T161 (Fig. 5D (circled)), which stabilises the P1–P1’ backbone in a “wrong” (proteolytically inactive) orientation.

Perhaps most importantly, in both the 3K24 crystal structure and our predicted dockings, the P2 sidechain lies close to residue 163, which is Gly in FhCatL5. The (polar) G163 backbone is exposed to the S2 subsite, and is likely to tolerate the proximity of Asp (Fig. 5A and B). In contrast, position 163 is Ala in FhCatL2; thus, the hydrophobic A163 sidechain protrudes into the S2 subsite (Fig. 5C (circled) and 5D), and therefore reduces the tolerance for hydrophilic (especially charged) ligand residues. This may also partially explain the general preferred acceptance of hydrophobic residues at P2 in FhCatL2. In order to further determine the influence of G163, we have performed 3000 dockings of the tetrapeptide substrate to the mutant FhCatL5 G163A. The predicted dockings are qualitatively similar to those obtained for FhCatL2 discussed above. In particular, while several predicted binding conformations closely resemble that of the H3 peptide in the 3K24 co-crystal structure, none of the dockings predicted the correct P1–P1’ backbone orientation. Thus, computational modelling suggests that G163 appears to be essential for binding of the target peptide in a proteolytically active conformation in FhCatL5, and is likely to be a key molecular determinant for P2 Asp acceptance in Fasciola.
cathepsin L. However, other determinants need to be taken into account to explain P2 Asp acceptance (or lack thereof) when HuCatL is taken into account, as discussed below.

4. Discussion

The S2 subsite exerts a dominant influence on the specificity of cysteine proteases, and in this study FhCatL5 showed a clear preference for Leu residues at P2. This preference for aliphatic over aromatic residues at the P2 position has also been reported for FhCatL1 and FhCatL2 [14,15], several other parasite cysteine proteases, and HuCatK and HuCatS [19,33,34]. This contrasts with the S2 subsite preference of HuCatL which more readily accommodates aromatic Phe residues over the aliphatic Leu residues. Differences at aa209 have been shown to contribute to this dramatic shift in specificity [14,35]. The PS-SCL also confirmed the substrate kinetics undertaken here, and in previous studies, which show FhCatL5 does not readily accommodate Pro residues in the S2 subsite [13].

The unusual ability of FhCatL2 to accommodate Pro at P2 has been previously reported [13–15], and recently FgCatL1G and FhCatL3 were shown to also have a preference for Pro at the P2 position in substrates [12,17,29]. These abilities have been partially attributed to the Y69 residue for FhCatL2 and W69 for FgCatL1G and FhCatL3. Smooker and colleagues (2000) were able to show that a L69Y substitution in FhCatL5 resulted in an increased ability to accommodate Pro at P2; a L69W substitution resulted in a similar outcome [12,13]. A Y69L substitution was introduced into FhCatL2. This is the reciprocal of the mutation created in the original study. This resulted in a decrease in the enzyme’s ability to accommodate Pro in comparison with the wildtype enzyme, with both the $K_m$ and $k_{cat}$ values being diminished. Similar to the corresponding L69Y substitution in FhCatL5, the change in specificity was only partial, once again indicating the importance of other residues in determining S2 subsite specificity. Along with an increase in the efficiency against Pro, the L69Y substitution in FhCatL5 also altered the respective preference for aliphatic and aromatic P2 residues. The L69Y substitution resulted in a large drop in enzyme efficiency towards Leu and Val at the S2 subsite when compared to Phe, thus producing more even substrate specificity. The alternate Y69L substitution in FhCatL2 resulted in the opposite scenario, with Leu and Val more readily accepted compared to Phe.

Stack and colleagues have shown that a L209A substitution in FhCatL1 increases the S2 subsite preference for aromatic residues (Phe, Trp, Tyr), while diminishing preferences for aliphatic (Leu, Val, Ile) residues [14]. The data in this study suggests that aa69 (while not having the same influence as aa209) can also play a role in the relative preference of the S2 subsite for aliphatic and aromatic residues.

Initial phylogenetic analysis of the Fasciola cathepsins differentiated the proteases into 4 distinct clades [9]. By virtue of codon usage it was suggested that Fasciola cathepsins with F69, C69 or S69 may have evolved during the evolution of the gene family. Transcripts for several F69 cathepsin L proteases have since been isolated from both Fasciola species. These cathepsins form their own clade and appear to be juvenile specific [6]. The sidechain of Phe is similar to Tyr, with the exception of lacking an –OH group. The L69F substitution in FhCatL5 had a similar influence to the L69Y substitution, although the increase in acceptance of Pro residues, and shifting of preference for aromatic and aliphatic residues was less dramatic than seen with the L69Y mutation. While the sizes of
the respective side chains are similar, the polar nature of Tyr compared to the non-polar Phe appears to be important to the S2–P2 interactions. In the native Fasciola cathepsins, the presence of the large F69 sidechain reduces the pocket size and interaction with the Pro ring in a manner similar to that proposed with Y69 would be expected to increase the ability to cleave substrates with P2 Pro residue. Combined with the enzymatic preferences for FgCatL1G [12], this supports the proposal that in the context of the phylogenetic tree, the “ancestral” enzyme had the ability to accept Pro, and the reduction in such activity observed with FhCatL1 and FhCatL5 is a recent occurrence [9].

The L69C and L69S variants produced in this study would possess large active sites due to the smaller size of the S69 and C69 side chains. These enzymes were able to be activated and could cleave natural substrates and synthetic peptides. While functionally stable, these variants had low overall efficiency (kcat/Km) against most of the synthetic substrates analysed, including those containing P2 Pro and Phe residues. The affinity of their S2 active sites for Z-FR-AMC was higher, with Km values about half those of FhCatL5; however, poor turnover meant the overall efficiency was reduced. Thus the lack of any such variants amongst the Fasciola cathepsin L proteases is probably due to the fact that there would be no functional advantage in expressing these proteases.

The PS-SCL data for FhCatL5 confirms the overall similar substrate specificity between the adult expressed Fasciola cathepsins, while also highlighting important differences. While FhCatL5 certainly shows significant overlap of substrate specificity for the P1–P3 subsites analysed when compared to FhCatL1 and FhCatL2, it also possesses distinct subsite preferences. Possibly the most important of these is the ability to accept P2 Asp, due to the profound influence the S2 subsite has on enzyme specificity. To acquire insights into the molecular basis for this relatively peculiar amino acid specificity amongst cathepsin L-like proteases, we have employed computational automated dockings to identify interactions between a tetrapeptide ligand and the enzymes FhCatL2, FhCatL5 and the mutant FhCatL5 G163A. Comparisons of the predicted dockings with a recent co-crystal structure of histone H3 peptide and HuCatL enabled determination of the extent to which

Fig. 5. Predicted dockings most closely resembling the HuCatL crystal structure for the tetrapeptide bound to FhCatL5 (side view (A) and top view (B)); and FhCatL2 (side view (C) and top view (D)). The ligand and residues of interest are shown in liquorice representation, with selected atoms colour coded as follows: cyan = carbon, blue = nitrogen, red = oxygen, and white = hydrogen. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
the cathepsin subtypes are capable of binding P2 Asp-containing peptides in proteolytically active conformations. Based on these analyses, we advance two hypotheses regarding the principal molecular determinants of P2 Asp acceptance. Firstly, for FhCatL5, the P1–P1′ backbone is oriented in a manner permissive to proteolytic cleavage. Secondly, the P2 sidechain lies close to residue 163 for both FhCatL5 and FhCatL2. Position 163 is occupied by a Gly in FhCatL5. The lack of a sidechain for G163 results in full exposure of its polar backbone to the S2 subsite, which is therefore favourable to interactions with Asp. In contrast, position 163 is occupied by an Ala in FhCatL2. Thus, for FhCatL2, the hydrophobic sidechain protrudes into the S2 subsite, reducing the tolerance for charged residues such as Asp. This may also partially explain the general preferred acceptance of hydrophobic residues at P2 for cathepsins with A163. While G163 is likely to be a key determinant for P2 Asp acceptance in the Fasciola cathepsins, we note that HuCatL also possesses G163, although it does not exhibit significant capability to accept P2 Asp residues [19]. However, inspection of the 3K24 co-crystal structure reveals an Asp at position 161 located in close proximity to the P2 sidechain position of the H3 peptide ligand. This is illustrated in Fig. 4B. For HuCatL, the presence of an acidic sidechain in position 161 may result in substantial electrostatic repulsion with the P2 Asp of a potential peptide/protein ligand. In contrast, position 161 is occupied by a (polar, neutral) Asn in FhCatL5, which will therefore not exhibit such electrostatic repulsion against P2 Asp (furthermore, all Fasciola cathepsin L subtypes exhibit either N161 or T161). Thus, we propose that the key requirements for P2 Asp acceptance in Fasciola and HuCatL are: 1) G163 and 2) a non-acidic residue at position 161. Fasciola cysteine proteases have been implicated in a range of functions. The ability of the adult protease FhCatL2 to digest collagen, facilitated by acceptance of P2 Pro, implicates it in host tissue migration [14], while the temporal expression profiles of FgCatL1G and FhCatL3, combined with their ability to cleave structural proteins, indicates a role in encystment and/or migration [6,12,17,29]. The observed unusual P2 Asp acceptance, coupled with the relatively low percentage of total adult cathepsin L secreted (5%) [36], strongly suggests a specific role for FhCatL5, rather than just redundancy or immunogenic variability. Other proteases that lie in the same clade as FhCatL5 also possess G163 (Supplementary Table S1). Previous work in our laboratory has shown that FhCatL5 is able to digest a range of host substrates including immunoglobulins, haemoglobin, laminin and fibrinogen [12]. While these substrates are also digested by FhCatL1 and FhCatL2, the preference of FhCatL5 for P2 Asp suggests the likelihood of other unique substrates. While acceptance of P2 Asp has been attributed previously to HuCatF [19] and a T. brucei cathepsin B [30], it has not been related to a specific substrate or function in either. The F. hepatica helminth defence molecule is found in adult fluke excretory/secretory material and has been proposed to be cleaved by Fasciola cathepsin L proteases in vivo to release bioactive peptides [29]. Interestingly, a putative cleavage site identified possessed P1–P3 residues of RDR, which corresponds highly with the subsite preferences of FhCatL5. Fasciola cysteine proteases are considered good drug and vaccine targets [4,37]. However, the development of an inhibitor needs to be selective to the Fasciola cathepsin while minimally interacting with host cysteine proteases. The successful identification of the rare ability to accept P2 Asp in FhCatL5 provides a possible direction for inhibitor design using this ability to selectively target this Fasciola cathepsin. In summary, at aa69 the introduction of large aromatic residues (Tyr, Trp, Phe) increases Pro P1′ accessibility by reducing subsite size and generating subsequent interaction with the P2 substrate. The relative preference for aromatic over aliphatic residues also increases, although this is countered by the smaller subsite, and steric hindrances caused by the size of the sidechain. The presence of L69 appears to increase the relative preference towards other aliphatic P2 residues, possibly through stabilising interactions in combination with those at aa209. The presence of the smaller S69 and C69 residues opens up the subsite; however, without the stabilising interactions provided by the side chains of the larger residues, the overall catalytic efficiency of these enzymes drops. Understanding the substrate preferences of the Fasciola cathepsins and what substitutions are required to bring about these preferences can lead to determination of their roles, and can provide a focus towards inhibitor design. The unique subsite preferences for FhCatL5 identified in this study, specifically acceptance of P2 Asp, indicates a unique functional role for FhCatL5.

Acknowledgements

The authors thank Prof. John Dalton for the FhCatL2 clone. We thank RMIT University for support. Charles S. Craik is funded by NIH grant CA 128765.

Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.biochi.2012.01.011.

References


